

ESTADÍSTIKA ETA DATUEN ANALISIA

**VI: Aldagai bakunaren deskribapen bateratua:
kaxa-diagrama. Heterogeneotasuna.**

Egilea: Josemari Sarasola



Gizapedia

gizapedia.hirusta.io

6.1 Kaxa-diagrama

6.1.1 Eraketa

6.1.2 Interpretazioa

6.2 Heterogeneotasuna

6.3 Ariketak

6. gaia: Aldagai bakunaren deskribapen bateratua: kaxa diagrama. Heterogeneotasuna.

6.1 Kaxa-diagrama

Kaxa-diagrama (ingelesez, *box plot*) zentroa, sakabanatzea eta beste ezaugarriak zehaztasunez azaltzen dituen diagrama sinple bat da. Horretaz gainera, *outlier* edo datu atipikoak aurkitzeko irizpide bat ematen du. John W. Tukey estatubatuar estatistikariak asmatu zuen 1969an, eta geroztik zabalkuntza handia izan du.

6.1.1 Eraketa

Pauso hauek jarraitu behar dira:

1. Lehenengo eta hirugarren kuartilak eta mediana kalkulatu behar dira.
2. Laukizuzen bat, kaxa alegia, marraztu behar da, ezkerretik lehenengo koartila eta eskuinetik hirugarren koartila mugatzat dituena.
3. Mediana goitik beherako marra batez irudikatzen da kaxa barnean.
4. Kuartil arteko ibiltartea kalkulatu behar da: hirugarren koartila ken lehenengo koartila.
5. $1.5(Q_3 - Q_1)$ kalkulatu eta kaxako alde banatara, lehenengo eta hirugarren kuartiletatik alegia, biboteak edo besoak luzatzen dira neurri hartan.

Luzapen hori behin behinekoa da.

6. Besoen luzapen horiek *luzapenak bukatu bitartean dauden azken datuetako balioetan* moztu behar dira. Balio horiek *albo-balioak* dira.
7. Biboteen muturretatik harago dauden datuak datu atipikotzat hartzen dira eta puntu banaz adierazten dira.

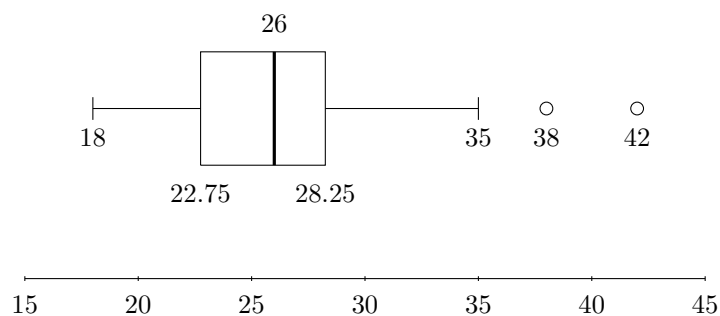
Adibidea: Autobus batek goizeko ibilbide bat egiteko behar dituen denborak jaso dira (minututan):

18,19,20,20,21,22,22,23,24,24,25,
 25,25,25,26,26,26,26,27,27,27,27,
 28,29,29,32,32,35,35,38,42

Kaxa diagrama eratu behar da.

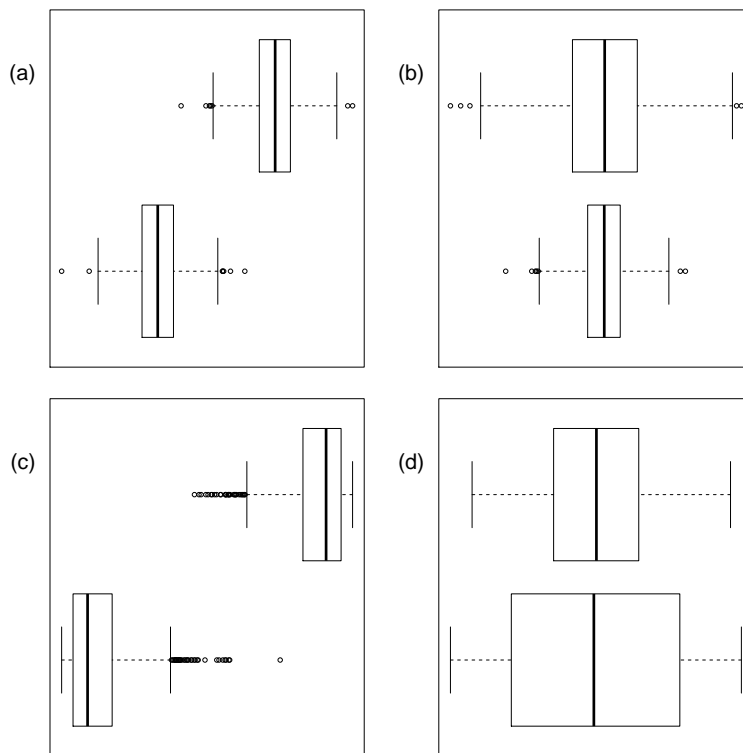
Kalkulu-taula (izarrarekin kaxa-diagrama eratzeko behar direnak)

$Me^{(*)}$	26
$Q_1^{(*)}$	22.75
$Q_3^{(*)}$	28.25
$1.5(Q_3 - Q_1)$	8.25
$Q_1 - 1.5(Q_3 - Q_1)$	14.5
$Q_3 + 1.5(Q_3 - Q_1)$	36.5
<i>albo-balioak azpitik</i> (*)	18
<i>albo-balioak gainetik</i> (*)	35
<i>outlier azpitik</i> (*)	-
<i>outlier gainetik</i> (*)	38, 42



6.1.2 Interpretazioa

- **Zentroari** buruz jakiteko, **medianari** erreparatu behar zaio;
- **sakabanatzeari** buruz, **kaxaren zabalera** ($IQR = Q_3 - Q_1$ neurriari, alegia) begiratu behar zaio;
- **alborapenari** buruz, kaxa barruko bi zatien neurriak alderatuko dira: ezkerrekoa eskubikoa baino zabalagoa bada, ezkererako alborapena dago; eskuinekoa zabalagoa bada, eskuinera izango da alboratua; eta biak berdintsuak badira, simetrikotzat har daiteke;
- **kurtosiari** buruz, kaxaren eta biboteen zabalera alderatuz dira: banakuntza leptokurtiko batean, biboteak estuak dira kaxaren aldean; oso banakuntza platikurtiko batean, bi biboteen zabalera kaxaren zabalera berdintsua da.



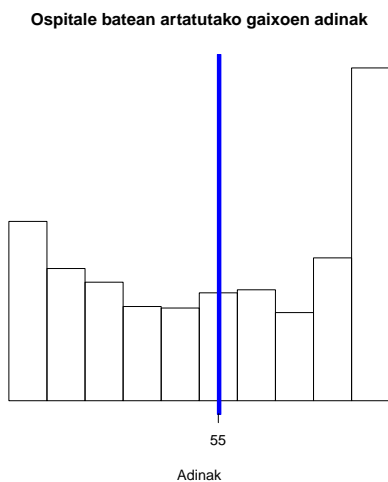
Irudia 6.1: (a) Zentro ezberdineko bi banakuntza, sakabanatze berdinekin. (b) Sakabanatze ezberdineko bi banakuntza, zentro berdinekin. (c) Ezkerrera (goian) eta eskuinera (behean) alboratutako banakuntzak. (d) Banakuntza leptokurtikoa (goian) eta platikurtikoa (behean)

6.2 Heterogeneotasuna

Heterogeneotasuna edo **ikerketa-heterogeneotasuna** ezaugarri ezberdinetako datu-multzoak batera jartzeagatik emaitzetan suertatzen den adierazgarritasun-eza da.

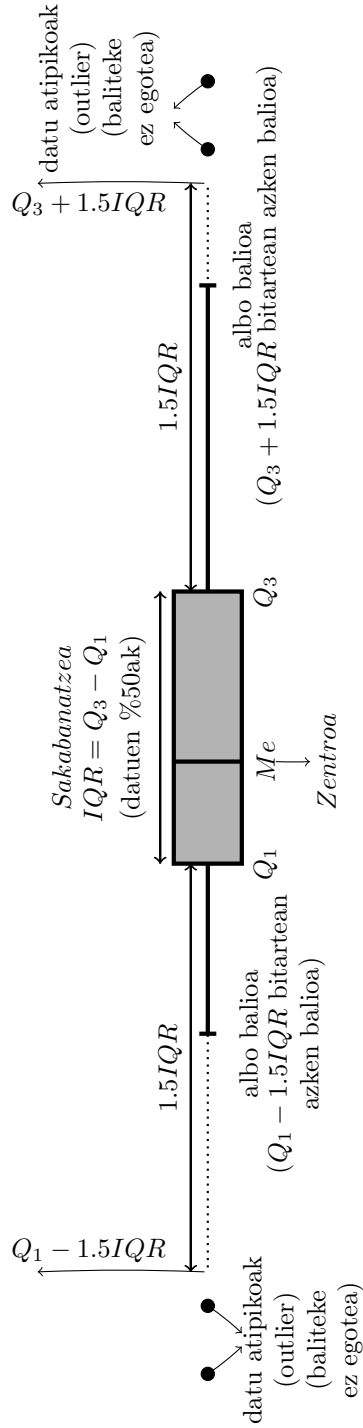
Adibide gisa, har ditzagun ospitale batean artatutako gaixoen adinak (ikus, irudia): ume jaioberriak eta haur txikiak badira dexente, gazteak eta helduak gutxiago, agureak, berriz, asko. Guztiak batera hartu eta batez besteko adina kalkulatzeko bada, errealitatearekin bat ez datorren emaitza batera helduko gara: haur txikien eta agureen adinak konpentsatuta, batezbestekoa 55 urtekoa da. Konklusio hori ez da egiazkoa, noski, eta haurrak, helduak eta agureak, ezaugarri ezberdinak dituzten taldeak, batera jarri direlako, heterogeneotasuna gertatu delako alegia, gertatu da.

Heterogeneotasuna gainditzeko, ezaugarri ezberdinak dituzten datu-multzoak bakoitza bere aldetik aztertu behar dira. Adibidean, jaioberriak, haurrak, gazteak eta helduak, eta zaharrak aparte aztertu beharko lirateke beraz, horien patologiak guztiz ezberdinak izaten direlako.



Bestela, moda aniztasuna, sakabanatze handia, alborapen handia eta U itxura izatea (adibidean bezalaxe) heterogeneotasunaren zantzu garbiak izaten dira.

Boxplot edo kaxa-diagramaren eskema



6.3 Ariketak

1. Enpresa bateko bi ostatuk hainbat hilabetetan izandako fakturazioak jaso dira (milaka eurotan):

A ostatua: 1.2-1.6-2.3-4.6-4.8-5-5.2-6.3-7-8-11.8-14.6

B ostatua: 3.4-5.2-6.3-6.6-6.8-7-7.2-7.4-7.8-8-8.1-8.3

Kaxa diagramak eratu eta interpretatu behar dira, zentroari eta sakabanatzeari buruz.

2. Bi autobusek ibilbide jakin bat egiteko 50 egunetan behar izan dituzten denborak jaso dira. Hona hemen datuok (minututan):

A autobusa

20 30 31 31 32 32 34 34 34 35 35 36 37 40 40 40 41
41 41 41 43 43 44 44 44 44 44 44 45 45 46 46 46 48
48 49 51 51 52 52 52 52 53 53 53 56 58 59 60 61

B autobusa

36 41 42 42 43 43 43 44 44 46 46 47 48 48 49 49 49
49 49 50 50 51 51 51 52 52 52 52 53 53 53 53 53 54
54 54 55 55 56 56 56 57 58 59 61 61 62 62 67 72

Kaxa-diagramak eratu grafiko berean eta interpretatu zentroari eta sakabanatzeari buruz. Zein da orokorrean autobus azkarrena? Zein autobus aukeratu behar da ordutegiak zehaztearren?

3. Datu-multzo batean hauek dira ezagunak:

- $Me = 40$
- $Q_1 = 36$ eta horren azpitik 6 datu daude.
- $Q_3 = 46$ eta horren gainetik 6 datu daude.

Asmatu itzazu lehen kuartilaren azpitik eta hirugarren kuartilaren gainetik dauden datuak,

- (a) ezkerreko bibotea zehazki erdira mozteko, eta ezkerrera bi datu atipiko uzteko;
- (b) eskuineko bibotearen luzera 4 unitate mozteko, eta eskuinera hiru datu atipiko uzteko.

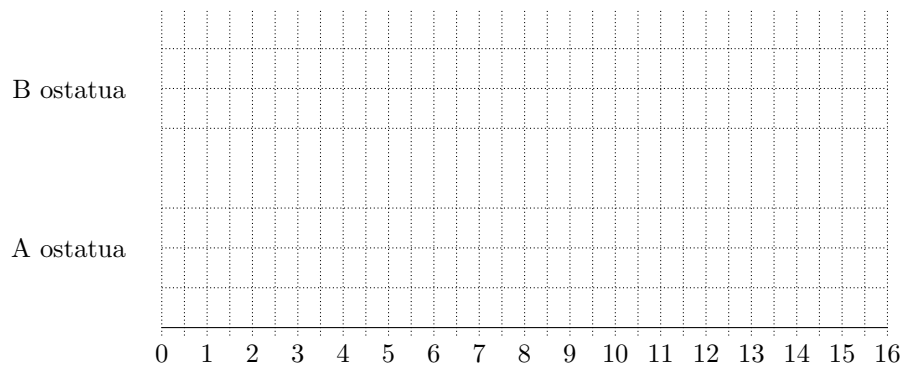
Azkenik, marraztu ezazu kaxa-diagrama.

Ebazpenak

1. ariketa

Kalkulu-taula

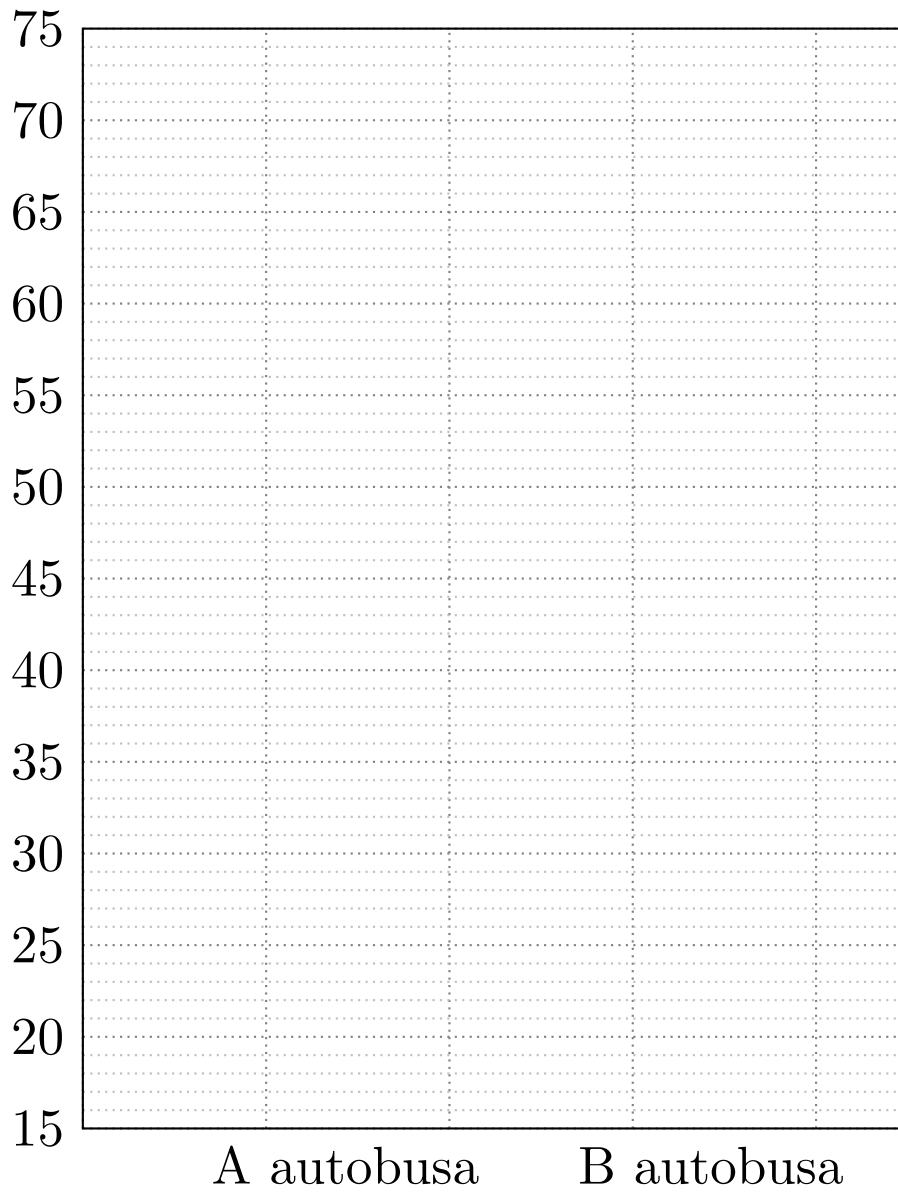
Parametroa	A ostaturia	B ostaturia
Me		
Q_1		
Q_3		
$1.5(Q_3 - Q_1)$		
$Q_1 - 1.5(Q_3 - Q_1)$		
$Q_3 + 1.5(Q_3 - Q_1)$		
<i>albo-balioak azpitik</i>		
<i>albo-balioak gainetik</i>		
<i>outlier azpitik</i>		
<i>outlier gainetik</i>		



2. ariketa

Kalkulu-taula

Parametroa	A autobusa	B autobusa
Me	25gn datua: 44	25gn datua: 52
Q_1	12.5gn datua: 36.5	12.5gn datua: 47.5
Q_3	37.5gn datua: 51	37.5gn datua: 55
$1.5(Q_3 - Q_1)$	21.75	11.25
$Q_1 - 1.5(Q_3 - Q_1)$	14.75	36.25
$Q_3 + 1.5(Q_3 - Q_1)$	72.75	66.25
<i>albo-balioak azpitik</i>		
<i>albo-balioak gainetik</i>		
<i>outlier azpitik</i>		
<i>outlier gainetik</i>		



3. ariketa

(a)

Printzipioz bibotearen luzera $1.5 \times (46 - 36) = 15$ da.

Lehen kuartilaren azpitik, besoa erdira mozteko, besoko azken datua $36 - 7.5 = 28.5$ izan behar da. Datu atipikoak $36 - 15 = 21$ baino txikiagoak izan behar dira. Beraz, lehen kuartilaren azpitik datuak hauek izan daitezke:

19-20-28.5-30-32-34

(b)

Hirugarren kuartilaren gainetik, besoari 4 unitate kentzeko, besoko azken datua $46 + 15 - 4 = 57$ izan behar da. Datu atipikoak $46 + 15 = 61$ baino handiagoak izan behar dira. Beraz, hirugarren kuartilaren gainetik datuak hauek izan daitezke:

50-52-57-62-63-64

